Stunting Classification Analysis for Toddlers in Bojongsoang: A Data-Driven Approach

Muhammad Ghiyaats Daffa[1], **Putu Harry Gunawan**[2]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[4]Divisi Digital Service PT Telekomunikasi Indonesia
[1]ghiyats@students.telkomuniversity.ac.id, [2]phgunawan@telkomuniversity.ac.id

Abstract

Stunting merupakan salah satu masalah kesehatan yang menjadi prioritas masalah kesehatan anak di Indonesia. Pencegahan stunting pada balita sangat diperlukan untuk menghindari dampak jangka panjang bagi balita dan masyarakat. Pencegahan stunting dapat dilakukan dengan memantau pertumbuhan balita. Oleh karena itu, dibutuhkan sebuah sistem yang dapat memprediksi kondisi stunting pada balita. Machine learning menawarkan banyak metode yang dapat digunakan untuk membangun sistem prediksi kondisi stunting pada balita. Penelitian ini menganalisis beberapa model machine learning yang berpotensi cocok untuk memprediksi kelas stunting, yaitu K-Nearest Neighbor (KNN), Random Forest (RF), dan Ensemble Learning yang disebut Boosted KNN (BK). Dataset yang digunakan dalam penelitian ini memiliki masalah ketidakseimbangan, dimana data stunting hanya sebesar 1% dari total dataset. Oleh karena itu, dilakukan oversampling pada dataset dengan cara membangkitkan dataset secara acak berdasarkan distribusi data yang tergolong dalam kelas minoritas. Hasil dari penjabaran oversampling ini terbukti memuaskan. Menerapkan data yang tidak seimbang memberikan rata-rata akurasi sebesar 98% untuk semua metode yang digunakan; namun, rata-rata makro skor F-1 terbukti tidak optimal untuk masing-masing metode, dengan 51,95% untuk KNN, 52,45% untuk RF, dan 53,55% untuk BK. Setelah data diseimbangkan dengan melakukan oversampling, rata-rata makro skor F-1 untuk semua metode meningkat secara substansial. Hasil yang baru adalah 93,55% untuk KNN, 97,70% untuk RF, dan 98,00% untuk BK, menggarisbawahi peran penting dalam mengatasi ketidakseimbangan data dalam meningkatkan akurasi prediksi.

Abstract

Stunting is one of the health problem priorities for children in Indonesia. Prevention of stunting in toddlers is needed to avoid the long-term effects for both the toddlers and the public. Stunting prevention can be done by monitoring the growth of toddlers. Therefore, a system that can predict stunting conditions in toddlers is needed. Machine learning offers many methods that can be used to build a system to predict stunting conditions in toddlers. This research analyzes some machine learning models that are potentially suitable to predict stunting classes, which are K-Nearest Neighbor (KNN), Random Forest (RF), and Ensemble Learning called Boosted KNN (BK). The dataset has an imbalance issue in this research, with the stunting data at only 1% of the total dataset. Therefore, oversampling of the dataset is done by generating a random dataset based on the distribution of the data that are classified as the minority class. The results of elaborating on this oversampling are shown to be satisfying. Applying imbalanced data gives an average of 98% accuracy for all methods used; however, the F-1 score macro average is shown not optimal for each of the methods, with 51.95% for KNN, 52.45% for RF, and 53.55% for BK. After the data is balanced by oversampling, the F-1 score macro average for all methods substantially increases. The new results were 93.55% for KNN, 97.70% for RF, and 98.00% for BK, underscoring the critical role of addressing data imbalance in improving predictive accuracy.

Keywords: stunting, machine learning, k-nearest neighbor, random forest, ensemble learning

1. Introduction

Stunting ranks among the foremost health concerns for children in Indonesia, denoting a condition where children under the age of five fail to thrive due to chronic malnutrition within the initial 1000 days of life, resulting in stunted growth [2, 17, 11]. Addressing this issue is crucial to circumvent enduring consequences for toddlers and the wider public, preventing long-term health decline [12]. Notably, President Joko Widodo has set a target to reduce stunting prevalence from 21% in 2022 to 14% in 2024, as outlined on the https://sehatnegeriku.kemkes.go.id website. Effective stunting prevention involves meticulously monitoring toddler growth, underscoring the need for an implementable system to predict stunting conditions.